



Research, part of a Special Feature on [Ecosystems and Society: Interactions Among Climate, Land Use, Ecosystem Services, and Society](#)

Comparing group deliberation to other forms of preference aggregation in valuing ecosystem services

[Mackenzie B. Murphy](#)¹, [Georgia Mavrommati](#)², [Varun Rao Mallampalli](#)³, [Richard B. Howarth](#)⁴ and [Mark E. Borsuk](#)³

ABSTRACT. Deliberative methods for valuing ecosystem services are hypothesized to yield group preferences that differ systematically from those that would be obtained through calculative aggregation of the preferences of participating individuals. We tested this hypothesis by comparing the group consensus results of structured deliberations against a variety of aggregation methods applied to individual participant preferences that were elicited both before and after the deliberations. Participants were also asked about their perceptions of the deliberative process, which we used to assess their ability to detect preference changes and identify the causes of any changes. For five of the seven groups tested, the group consensus results could not have been predicted from individual predeliberation preferences using any of the aggregation rules. However, individual postdeliberation preferences could be used to reconstruct the group preferences using consensual and rank-based aggregation rules. These results imply that the preferences of participants changed over the course of the deliberation and that the group preferences reflected a broad consensus on overall rankings rather than simply the pairwise preferences of the majority. Changes in individual preferences seem to have gone largely unnoticed by participants, as most stated that they did not believe their preferences had substantially changed. Most participants were satisfied with the outcome of the deliberation, and their degree of satisfaction was correlated with the feeling that their opinion was heard and that they had an influence on the outcome. Based on our results, group deliberation shows promise as a means of generating ecosystem service valuations that reflect a consensus opinion rather than simply a collection of personal preferences.

Key Words: *deliberative multicriteria evaluation; participatory methods; shared values; trade-off weights; value formation*

INTRODUCTION

The ecosystem services concept has traditionally represented a means of assigning monetary values to the services being provided by ecosystems. The idea is that if such services were monetized, then natural systems would be given greater consideration in government planning, policy, and projects (TEEB 2010). Of course, not all services provided by ecosystems are readily monetized, especially cultural services such as recreation, spiritual inspiration, intellectual stimulation, and aesthetic beauty. In theory, the relative value of such amenities could be quantified without reference to money using methods such as multicriteria decision analysis (Keeney and Raiffa 1993) or the Analytic Hierarchy Process (Saaty 1980). However, such methods were developed to characterize the values of an individual, and problems arise when attempting to aggregate the values of multiple individuals to achieve a “societal-level” valuation. Preference description using multiattribute value functions, for example, has a sound theoretical basis at the level of a single decision-maker (Keeney and Raiffa 1993), but there is no corresponding axiomatic foundation for social preferences. This is because individual welfare is a subjective condition that cannot be measured objectively or compared interpersonally (Robbins 1938). The corollary is that it is impossible to derive an aggregate measure of social welfare that satisfies a set of reasonable criteria for rationality and fairness (Arrow 1951). Given that most ecosystem service valuations start by eliciting the preferences of individuals (Hostmann et al. 2005, Koschke et al. 2012, Karjalainen et al. 2013), there is a need to better understand how personal values can be effectively translated into descriptions of social value.

Group deliberation is being increasingly explored as a means of aggregation of individual preferences through “mutual consent” rather than through calculation (Wilson and Howarth 2002, Howarth and Wilson 2006, Proctor and Drechsler 2006, Stagl 2006, Kenter et al. 2011, Liu et al. 2011, Zia et al. 2015, Lienhoop and Völker 2016). The basic idea is that small groups of citizen stakeholders are brought together to discuss and debate the relative importance of a particular set of public goods. This process forces participants to think critically about their own preferences and engage deeply with alternative perceptions of value (Irvine et al. 2016). The goal is to reach an informed group judgment based on widely held social values rather than simply on a collection of individual preferences (Wilson and Howarth 2002). This is expected to result in a broader community understanding of trade-offs and an increased likelihood of conflict resolution earlier in the policy-making and planning process (Mavrommati et al. 2016).

The deliberative approach to ecosystem service valuation also creates a forum in which subject matter experts can provide participants with additional information as needed about the economic or environmental systems under consideration. In fact, given the recognition that environmental values are likely to be socially constructed rather than extant in the minds of individuals, it is essential that the groups engage in learning and reflection as a social unit (Zografos and Howarth 2010). This opportunity for participants to form collective preferences may also encourage thoughtful inclusion of social equity and sustainability considerations (Kenter et al. 2015).

¹Dartmouth College, Hanover, NH, ²School for the Environment, University of Massachusetts, Boston, MA, ³Department of Civil and Environmental Engineering, Duke University, Durham, NC, ⁴Environmental Studies Program, Dartmouth College, Hanover, NH

While deliberative methods offer a compelling approach to ecosystem service valuation that may result in more thoughtful and socially relevant preference descriptions, some key questions remain regarding the relation between personal values, the deliberative process, and assessments of social value. In a deliberative multicriteria evaluation of recreation and tourism activities in the upper Goulburn-Broken Catchment of Victoria, Australia, participants' rankings of 13 ecosystem services diverged substantially in a predeliberation questionnaire (Proctor and Drechsler 2006). After deliberation, their individual importance rankings were much more similar. They also reflected a shift from valuing recreational access to valuing conservation, thereby suggesting a greater appreciation of shared social values over individual values. This study did not elicit joint group preferences.

Employing a three-stage choice experiment to elicit the monetary values of coastal attributes in the Firth of Forth in central Scotland, Kenter et al. (2016) found that there was a substantial decrease in the value placed on ecosystem services over the course of the three rounds. Predeliberation individual values were the highest, followed by postdeliberation individual values; the deliberated group values were the lowest. The author suggests that this pattern is the result of the participants becoming more aware of the implications of their choices, and thus less inclined to place high values on environmental attributes merely as an expression of their support for conservation.

In eliciting ecosystem service values for proposed UK marine protected areas (MPAs), Kenter et al. (2016) compared five sets of values: individual values from an online survey, individual and group values following exchange and discussion of information on MPAs, and individual and group values following both exchange and discussion of information and exchange of experiences through storytelling. Both sets of deliberated group values were significantly lower than individual values, as assessed in both the online survey and after the deliberations. Like Kenter (2016), the authors suggest that this is because participants developed more clearly formed beliefs during deliberation, specifically around issues such as access and site restrictions. Indeed, participants were more confident about their deliberative group values and believed they would be more appropriate to use for decision-making. Group values also more closely corresponded to measures of subjective well-being, thereby suggesting they were a better representation of true preferences than were individual values.

Each of the studies described in this section implicitly or explicitly aggregated individual values by calculating the mean before comparing against deliberative group values. Indeed, this practice of averaging or summing the preferences of individuals is widespread in ecosystem service valuation (Wegner and Pascual 2011, Meinard et al. 2016), whether the preference information is expressed in cardinal form, as in the studies described, or in ordinal form (e.g., Haikowicz 2006, García-Llorente et al. 2012, Haida et al. 2016, Zoderer et al. 2016). However, as implied by Arrow's (1951) Impossibility Theorem, using the mean to aggregate individual preferences does not hold a place of privilege over other methods of aggregation. Other methods can generate different aggregated preferences, which may compare differently against deliberative group preferences. Because each method

conforms with different social choice criteria, these comparisons may provide some insight into the implicit conditions employed by the groups to reach their deliberative result.

These observations motivate the following research questions:

1. Do deliberative group preferences differ systematically from those that would be obtained through calculative aggregation of the preferences of the participating individuals?
2. Does the relationship between deliberative group and aggregated individual preferences depend on the aggregation method used? If so, what does this observation tell us about the nature of the group deliberative process?
3. Do participants' preferences change as a result of the deliberative process? If so, do participant preferences tend to converge?

These questions were addressed using data from seven panels of citizen stakeholders who were tasked with assessing the relative value of 10 different ecosystem services being provided by the upper Merrimack River watershed, New Hampshire (NH).

METHODS

Study location

The upper Merrimack River watershed, defined by a point just south of Manchester, NH, has an area of 8000 km² and a population of 410,000. Forest is currently the dominant land cover, but the region is experiencing rapid population growth and associated land cover change. This is leading to increased water use, nitrogen discharge, and other environmental impacts associated with development. In addition, climate change is anticipated to lead to warmer overall temperatures and greater and more variable precipitation. Impacts on the region's natural amenities and the resulting effects on winter and summer tourism are likely to have important economic consequences.

As part of a larger effort to assess the impacts of changing climate and land use on the provision of ecosystem services in the state, we held four full-day workshops with residents of the upper Merrimack watershed. The goal of each workshop was for participants to assess the relative importance of 10 preselected ecosystem services to future residents of the watershed, given a specific future socioeconomic scenario for the year 2100. Details of participant recruitment and selection, scenario presentation, and full valuation results are reported by Mavrommati et al. (2016). We describe only the components of the study that are essential to addressing the research questions posed in the Introduction.

Workshop format and survey questions

Participant recruitment by local print and online media yielded 286 respondents, 217 of whom lived in the upper Merrimack watershed and were available to participate in one of our workshops. From among these 217, we invited 96 who, to the greatest degree possible, represented the demographics of the NH population in terms of age, sex, income, and political affiliation. A total of 67 participants attended the workshops, which were held on four separate dates in September 2015. Based on the number of attendees at each workshop, we formed two groups on the first date and three on each of the remaining dates. Each group

Table 1. Ecosystem services evaluated by participants.

Domain	Ecosystem service name	Definition	Units
Land	Farm Land	Total area of agricultural land (both cropland and pasture) divided by the population	Acres per person
	Forest Cover	% of total watershed area that is forest	% of total land area
Climate	Forest Type	% of forest that is suitable for maple trees	% of forest
	Heat Regulation	Days per year with temperature > 90° F	Days
	Snow Cover	Days per year with snow > 6 inches	Days
Water	Recreation Days	Days per year with temperature between 70° and 90° F	Days
	Fish Habitat	Total upstream river length and duration impaired by temperature, chloride, or discharge	% river miles
	Coastal Health	Nitrogen export to estuary exceeding regulatory threshold	Tonnes nitrogen per year
	Water Supply	Population duration of water supply stress	Million person-days
	Flood Protection	Population duration of potential flood impact	Thousand person-days

was comprised of five to seven participants to achieve a comparably diverse mix of demographic characteristics. For their involvement, participants received coffee and pastries, lunch, travel cost reimbursement, and \$100.

Participants spent the morning of each workshop being introduced to the deliberative process, the ecosystem services concept, and the nature of their valuation task. A specific future socioeconomic scenario and definitions of the 10 selected ecosystem services were then presented by project members. In the afternoon, participants were divided into groups, and each group was led to a separate room by a professional facilitator. After a warm-up exercise, each group then performed the choice task described by Mavrommati et al. (2016). This included a determination of the relative importance of three or four ecosystem services within each of three domains: in turn, land, climate, and water (Table 1).

For each domain, groups were given three or four cards, each representing a hypothetical future state of the world with dashboard-style infographics indicating the level of provision of each ecosystem service in that domain. On each card, one ecosystem service was indicated at its worst possible level, and the other two (or three) were indicated at their best possible levels, with worst and best levels determined by a separate modeling exercise (Samal et al. 2017). The facilitator then asked each participant to share with the rest of the group his or her thoughts and arguments for a particular preference ordering of the states of the world represented by the cards.

Next, the group was given a measurement stick scaled from 0 to 100 and was told that the high end should be interpreted as their shared degree of preference for a state of the world in which all ecosystem services are at their greatest possible level and the low end should be interpreted as their degree of preference for the state of the world that is least preferred among the cards they were given. The group was then asked to place the cards along the measurement stick, with location and spacing representing their consensual relative preferences for the states of the world represented by the cards. A scientist from each domain was available to answer participant questions about the ecosystem services as they arose. Scientists were asked not to express their personal opinions or value judgments.

Deliberation, discussion, and debate continued within each group until the participants were able to agree on a final positioning of cards. Facilitators managed the process so that each person was able to contribute to the discourse and was finally able to agree verbally with the group's decision. Time was managed so that each domain was considered for approximately one hour. Facilitators documented the final positions of the cards in each domain. The importance ranking of each ecosystem service, compared with others in its domain, was then determined by the relative position of the card on which that particular ecosystem service was indicated at its worst possible level (Schuwirth et al. 2012). While numerical trade-off weights could also be determined by the exact placement of the cards on the measurement stick (Mavrommati et al. 2016), we were concerned only with the relative ordering.

All participants completed individual surveys both before and after the group deliberations, and were asked to rank the relative importance of the ecosystem services in each domain. The standard deviation was used to characterize the degree of disagreement in rankings, between participants in a group and between groups. Higher values indicated greater disagreement.

At the end of the day, participants were asked additional questions about their experience (each with possible responses consisting of "Extremely," "Considerably," "Moderately," "Slightly," and "Not at all"):

1. How much did your opinions about the relative importance of ecosystem services change over the course of the workshop?
2. How well do you feel that your opinion was heard during the group deliberation?
3. How influential were you on the outcome of the group deliberation?
4. How influential were the scientists on the outcome of the group deliberation?
5. How satisfied were you with the outcome of the group deliberation?

To assess the determinants of participants' satisfaction with the deliberative process and their changes in preferences, we calculated the Pearson cross-correlations (ρ) between respondent

answers to these questions, as well as the Pearson correlations with the following four metrics: the correlation between each participant's pre-deliberation and post-deliberation rankings (τ_{ind}), the correlation between each participant's pre-deliberation rankings and their group's deliberative group ranking (τ_{pre}), the correlation between each participant's post-deliberation rankings and their group's deliberative group ranking (τ_{post}), and the difference between τ_{post} and τ_{pre} ($\tau_{post} - \tau_{pre}$). This last measure is used as an indication of the degree of convergence of a participant's ranking toward the group ranking from pre-deliberation to post-deliberation. All correlations were calculated as the Kendall rank correlation coefficient, τ , which measures the ordinal association between two quantities. The value of τ will be near 1 when two rankings are similar, near 0 when they are unrelated, and near -1 when they are conflicting.

Aggregations based on social choice theory

Deliberative methods of valuation were developed in part to overcome Arrow's Impossibility Theorem, which states that no mathematical or logical rule exists for aggregating individual preference orderings into a joint, or social, preference ordering while also satisfying conditions of monotonicity, nondictatorship, independence of irrelevant alternatives, individual sovereignty, and universality (Arrow 1951). Any possible calculative aggregation rule must therefore violate at least one of these criteria. Nevertheless, many aggregate ranking rules have been proposed which conform with various combinations of these, or other, criteria. Comparing the results of these aggregation rules when applied to individual rankings against deliberative group rankings may provide some indication of how the groups reached their deliberative result. We chose five aggregation methods as the basis for our comparisons, the key characteristics of which we summarize here and in Table 2. Detailed descriptions of the methods in the context of environmental and conservation planning are provided by Burgman et al. (2014).

Table 2. The criteria satisfied by each of the considered aggregation methods.

Criterion	Aggregation method				
	Plurality	Borda	Hare	Copeland	Kemeny
Arrow criteria					
Monotonicity	1	1	0	1	1
Nondictatorship	1	1	1	1	1
Indep. of irrelevant alt.	0	0	0	0	0
Individual sovereignty	1	1	1	1	1
Universality	1	1	1	1	1
Pareto efficiency	1	1	1	0	1
Condorcet winner	0	0	0	1	1
Majority winner	1	0	1	1	1

Extended Plurality: Only the first-ranked choice for each participant is considered, and the choice with the most first ranks is considered to be the aggregate first rank (Chamberlin 1985). The choice with the second-most first ranks is the aggregate second rank, and so on. Choices with an equal number of first ranks are considered tied. The argument for plurality ranking is that it adheres to the "one person, one vote" principle in which

each individual is able to indicate only his or her most preferred choice. It also satisfies the criterion of Pareto efficiency, which states that if every individual prefers ranks that are one option higher than another, then so must the resulting aggregate ranking. The plurality method, however, violates the Condorcet criterion, which states that an aggregate ranking should have the property that the choice that is most preferred by most participants in all possible pairings against the other choices is ranked first, and the choice that is least preferred by most participants in all possible pairings is ranked last, and that this holds recursively for the intermediate choices. It also violates the independence of irrelevant alternatives criterion, which states that the aggregate relative ranking of two choices A and B should depend only on the individual participant preferences between A and B and should not be influenced by consideration of an additional choice C. This criterion is violated by many aggregation systems, including all of those considered here.

Extended Borda count: Choices accrue points for each ranking position they receive, such that for n choices, first position is worth $n - 1$ points, second position is worth $n - 2$ points, and so on (Chamberlin 1985). After all points are summed for each choice, the choices are ranked according to their point values to achieve the aggregate ranking. Choices with equal point values are tied. Because the Borda count gives substantial consideration to a participant's lower ranked choices, it tends to support rankings that are supported by a broad consensus among participants rather than necessarily the ranking of a majority. In fact, the Borda count is the only one of the methods we considered that violates the majority criterion (Table 2), which states that if one choice is highest ranked by most individuals, then that choice must be the most preferred in the aggregate ranking. The Borda ranking also violates the Condorcet criterion.

Extended Hare: The choice that has the least number of first ranks is eliminated from all participant rankings such that all first rank selections are replaced with the next higher ranked remaining choice (Chamberlin 1985). This continues until only one choice remains or the remaining choices are tied. Choices are then ranked in reverse order of elimination. Like Plurality and Borda, the Hare method violates the Condorcet criterion. It also is the only method that violates the monotonicity criterion. A ranking system is monotonic if it is not possible for a choice to be ranked higher in the aggregate ranking as a result of some individuals lowering their ranking, and vice versa (while no other rankings are altered).

Kemeny rule: A matrix is first created that counts pairwise preferences of participants. A score is then calculated for all possible rankings, which equals the sum of the pairwise counts that apply to that ranking. The ranking that has the largest score is then chosen as the aggregate ranking (Kemeny 1959). This ranking is equivalent to the one that minimizes the sum of the Kendall tau distances to the individual participants' rankings. The Kemeny rule satisfies the Condorcet criterion as well as the criterion of Pareto efficiency.

Copeland's method: Also referred to as the pairwise comparison method, this method ranks choices by their total number of victories in pairwise preference comparisons (Copeland 1951). This method is easy to calculate, easily understood, and, notably, satisfies the Condorcet criterion. However, it is the only method we considered that violates the criterion of Pareto efficiency.

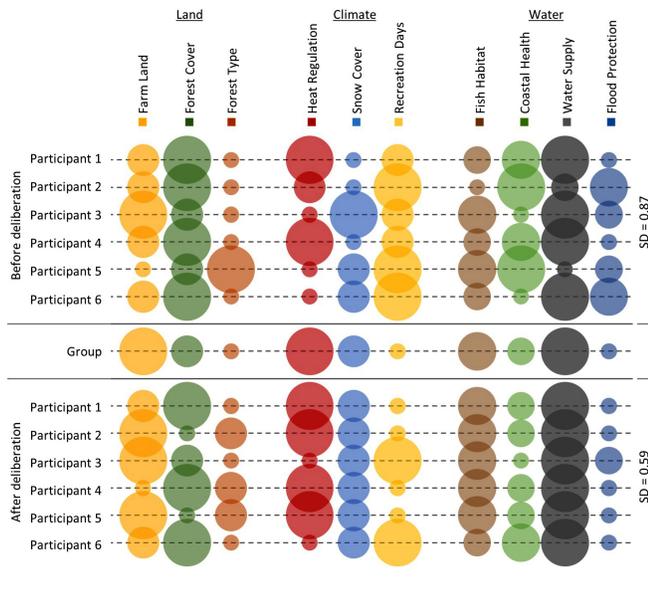
Notably, Copeland’s method is consistent with the Analytic Hierarchy Process (Saaty 1980).

For each group, we applied each of these five aggregation rules to the rankings provided by that group’s participants both before and after the group deliberations. We then used the Kendall rank correlation coefficient, τ , to compare each of these rankings against the group deliberative outcomes.

RESULTS

Of the 11 groups that participated in our workshops, one was not able to reach consensus for all ecosystem service domains, and three did not have individual ranking data collected. Therefore, our analysis concerns data from seven groups. Fig. 1 shows the form of the data for one of these groups, which consist of the rankings of ecosystem services within each of three domains as provided by (i) individual participants before group deliberation, (ii) the group’s deliberative result, and (iii) individual participants after group deliberation.

Fig. 1. Full ranking data from Group 7, corresponding to individual participants before group deliberation (top), the group’s deliberative outcome (middle), and individual participants after group deliberation (bottom). Size of the circles indicates the importance ranking, with large circles being most important. Standard deviations (SD) indicating the degree of disagreement between participant rankings averaged across domains are indicated in the right margin.



Individual rankings

Before deliberation, there was substantial disagreement among individuals in their ranking of ecosystem services, both overall and within each group, as indicated by high standard deviations (Table 3). Groups 1, 3, and 5 were particularly discordant. After deliberation, however, the individual rankings provided by each group’s participants became much more similar to one another. Standard deviations dropped after deliberation for all but Group 4. Standard deviations for individual ecosystem services (Table 4) show that this improved agreement among individuals within

groups occurred most prominently in the ranking of water services. In particular, individual participants’ rankings of Coastal Health and Water Supply showed the largest reduction in standard deviation over the course of the deliberation.

Table 3. Standard deviations between members of each group in ecosystem service rankings, averaged across ecosystem services.

Group	Before deliberation	After deliberation	Change
1	0.93	0.55	-0.37
2	0.63	0.52	-0.11
3	0.92	0.78	-0.14
4	0.86	0.91	0.05
5	0.91	0.67	-0.24
6	0.89	0.68	-0.21
7	0.87	0.59	-0.28
Average	0.86	0.67	-0.19

Table 4. Standard deviations between members of each group in ecosystem services rankings, averaged across groups.

Ecosystem service	Average within-group entropy			
	Before deliberation	After deliberation	Change	
Land	Farm land	0.81	0.63	-0.17
	Forest cover	0.78	0.63	-0.15
	Forest type	0.69	0.46	-0.23
Climate	Heat regulation	0.77	0.73	-0.03
	Snow days	0.82	0.62	-0.20
	Recreation days	0.91	0.75	-0.16
Water	Fish habitat	0.84	0.73	-0.11
	Coastal health	1.12	0.86	-0.27
	Water supply	1.06	0.74	-0.32
	Flood protection	0.78	0.57	-0.21

Deliberative process

Initially, participant discussions focused largely on their own experiences with the ecosystem services being discussed. However, most groups quickly turned their attention to understanding the specific properties of the ecosystem services and their interrelations, as well as the socioeconomic implications of changes in ecosystem services. This understanding was fostered through directed questions to the scientists. For example, Group 2 was initially divided with respect to the importance of Heat Regulation versus Snow Cover. This division was eventually resolved after a long discussion in which some participants emphasized the impacts of reduced snow cover on winter tourism and asked scientists about the implications for the seasonal hydrological cycle. Participants also discussed the ability of humans to adapt to rising temperatures and develop technological substitutes for heat regulation, such as air conditioning.

Values-based discussions were also concerned with the degree to which the loss of ecosystem services could be offset by substitutes and mitigation measures or reversed by future policies. Ecosystem service losses that were considered to be reversible or readily mitigated were ranked lower in importance. For example, after discussion, flood attenuation was given a lower ranking since

participants recognized readily available mitigation measures and believed that communities could recover from flood events. Loss of fish habitat or health of coastal zones was considered by most groups to be largely irreversible and without substitutes, which resulted in a greater importance ranking.

In their discussions, participants gave appropriate consideration to the magnitude of differences between the best and worst possible levels presented to them. For example, a participant in Group 1 was convinced by another participant that Coastal Health should receive a higher ranking than Water Supply because its indicator showed a much greater relative difference across scenarios. In placing their group's cards along the measurement stick, with location representing their relative preferences, all groups sought agreement through evidence-based argumentation and consensus building rather than by tallying votes, allocating points, or using other quantitative means. Most groups were very deliberate about placement of the cards, and even debated small differences in placement.

Deliberative rankings

Deliberative rankings of ecosystem services (Fig. 2) were broadly similar across groups. For the land domain, Farm Land was ranked most important by four of the groups, and Forest Type was ranked least important by four groups. However, Groups 4 and 5 ranked these in opposite order, which led to overall high across-group standard deviations for these services. In the climate domain, summertime Heat Regulation was ranked most important by five groups and second most important by the other two. Recreation Days was ranked least important by all groups. The climate services had concomitantly low across-group standard deviations. In the water domain, there was more disagreement, especially about the relative importance of Water Supply. The across-group standard deviation of 1.36 was highest for this ecosystem service. Notably, Group 4 was the only group that assigned top importance to two ecosystem services in two separate domains (Fig. 2). This was also the only group for which the overall standard deviation between participants increased following deliberation (Table 3).

Aggregated individual rankings

None of the considered aggregation rules, as applied to the predeliberation individual rankings, could have accurately predicted the deliberative group rankings (Fig. 3; Table 5). All methods had average Kendall tau values that were less than 0.3. For some of the groups, many of the methods had negative values, which means that there was more disagreement than agreement between calculative aggregated and deliberative group rankings. The Borda method performed best overall, with the highest overall average agreement of 0.26 and the highest agreement among the various rules in five out the seven groups. The correspondence between aggregated and deliberative rankings differed greatly across groups, with a high correspondence for Groups 1 and 6 and a very low correspondence for Groups 2, 3, and 5.

When applied to the individual postdeliberation preference rankings, some of the considered aggregation rules reproduced the deliberative group results quite well (Fig. 3; Table 5). The Borda, Kemeny, and Copeland methods performed best overall, with high Kendall tau values for most groups, including some perfect correlations. It is important to note that this was not

because all participants in those groups were in agreement after the deliberation (Table 4).

Fig. 2. Group deliberative rankings of ecosystem services. Size of the circles indicates the importance ranking, with large circles being most important. Asterisks indicate ties. Standard deviations (SD) indicating the degree of disagreement between groups for each ecosystem service are indicated in the bottom margin.

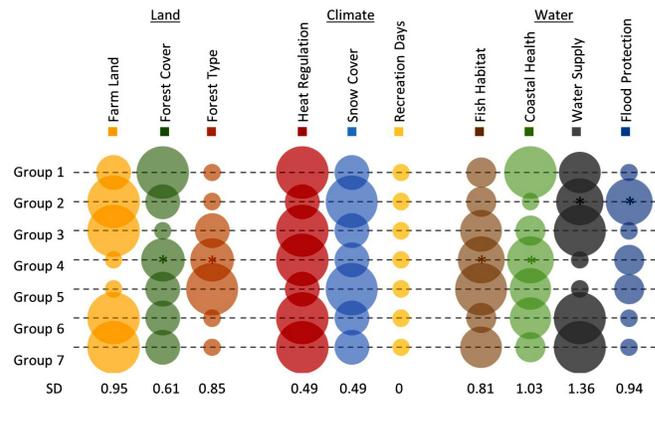
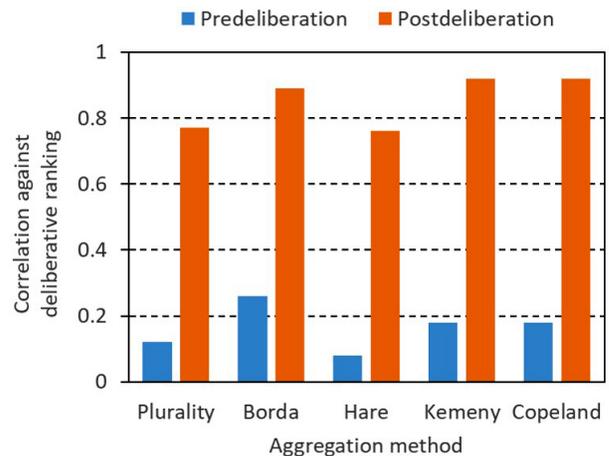


Fig. 3. Bar plot of Kendall correlations between deliberative group rankings and aggregated individual rankings assessed before and after the deliberation. Correlations shown are the averages across groups.



Individual rankings surveyed after deliberation were much more similar to the group deliberative rankings than those surveyed before deliberation, as gauged by the group-average values of the Kendall correlation coefficients (Table 6). This indicates that individual preferences were changed by the deliberative process in a manner that brought them each closer to the deliberative ranking.

Participant perceptions

Questions about participants' experiences revealed that participants were generally unaware that their preferences changed over the course of the deliberation process: almost 60%

Table 5. Kendall correlations between aggregated individual rankings and deliberative rankings for each group.

Aggregation rule	Group							
	1	2	3	4	5	6	7	Avg.
Before deliberation								
Plurality	0.63	-0.14	-0.33	-0.06	-0.07	0.67	0.12	0.12
Borda	1.00	-0.07	-0.09	0.32	-0.30	0.71	0.27	0.26
Hare	0.63	-0.14	-0.47	0.02	-0.27	0.67	0.12	0.08
Kemeny	0.94	-0.13	-0.09	0.22	-0.53	0.74	0.12	0.18
Copeland	0.94	-0.13	-0.09	0.22	-0.53	0.74	0.12	0.18
After deliberation								
Plurality	0.97	0.86	0.82	0.66	0.77	0.71	0.63	0.77
Borda	1.00	0.97	0.91	0.82	0.80	0.77	1.00	0.89
Hare	0.97	0.86	0.71	0.66	0.77	0.71	0.63	0.76
Kemeny	1.00	0.97	0.86	0.91	0.80	0.94	0.94	0.92
Copeland	1.00	0.97	0.89	0.91	0.80	0.94	0.94	0.92

answered “Slightly” or “Not at all” (Fig. 4a). Further, there was only a weak negative correlation ($\rho = -0.10$), with τ_{ind} , a measure of the actual consistency between their individual pre-deliberation and postdeliberation rankings. Their self-assessment of change was most strongly correlated with $\tau_{post} - \tau_{pre}$, an indication of their degree of convergence to the group ranking from pre-deliberation to postdeliberation ($\rho = 0.36$), as well as their assessment of the influence of scientists on the outcome ($\rho = 0.30$) (Table 7).

Table 6. Kendall correlations between individual rankings and deliberative rankings, averaged by group.

Group	Before deliberation	After deliberation	Change
1	0.10	0.74	+0.64
2	0.13	0.70	+0.56
3	0.04	0.48	+0.43
4	0.11	0.29	+0.18
5	-0.04	0.64	+0.68
6	0.30	0.68	+0.38
7	0.08	0.56	+0.48
Avg.	0.10	0.58	+0.48

Participants generally felt that their opinion was heard during the deliberations, with 86% answering “Considerably” or “Extremely.” In addition to being strongly correlated with their perception of influence ($\rho = 0.35$) and their feeling of satisfaction ($\rho = 0.65$), this feeling also correlated with τ_{post} , the correspondence of their postdeliberation rankings with the group rankings. The perception of participants on the influence of the scientists on the outcome was strongly negatively correlated with both τ_{pre} ($\rho = -0.31$) and τ_{ind} ($\rho = -0.37$), and was positively correlated with $\tau_{post} - \tau_{pre}$ ($\rho = 0.36$).

Most participants were satisfied with the outcome of the deliberation (76% answered either “Considerably” or “Extremely”). The response to this question correlated strongly with whether they felt their opinion was heard ($\rho = 0.65$), with their perception of influence ($\rho = 0.22$), and with the correspondence of their postdeliberation rankings with the group rankings, τ_{post} ($\rho = 0.23$).

Fig. 4. Pie charts indicating participant responses to five questions about their experience that were asked after the deliberative process.

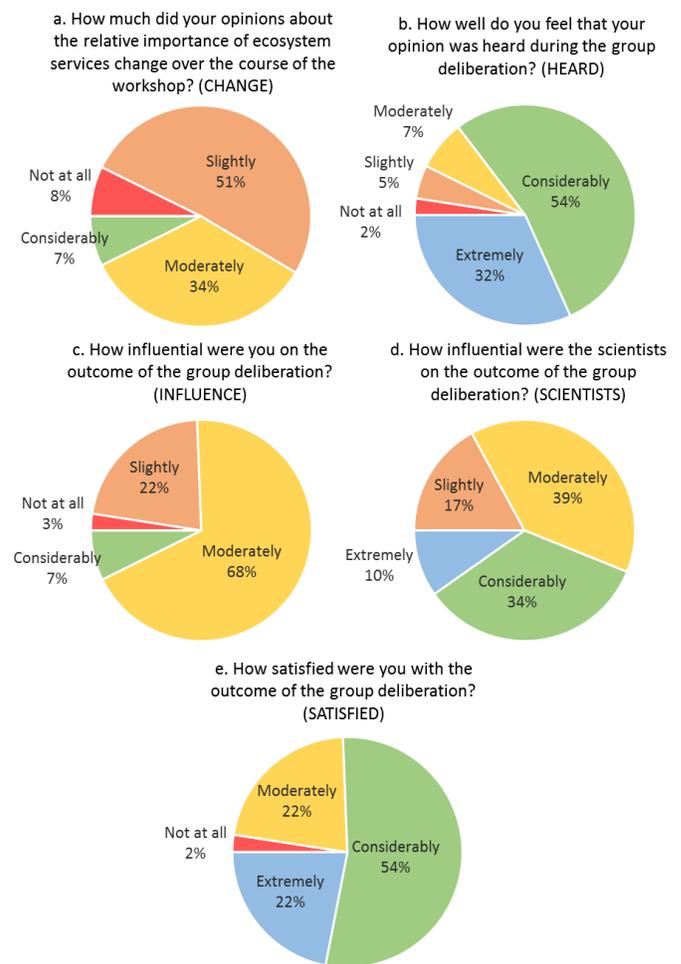


Table 7. Pearson correlations among respondent answers to questions about their experience and as compared to measures of changes in their preference rankings, τ_{ind} , τ_{pre} , τ_{post} , $\tau_{post} - \tau_{pre}$. Labels are defined in Fig. 4 and in the text.

	CHANGE	HEARD	INFLUENCE	SCIENTISTS	SATISFIED	τ_{ind}	τ_{pre}	τ_{post}	$\tau_{post} - \tau_{pre}$
CHANGE		-0.08	-0.04	0.30	0.09	-0.10	-0.19	0.26	0.36
HEARD	-0.08		0.35	-0.03	0.65	0.14	0.25	0.31	0.04
INFLUENCE	-0.04	0.35		-0.14	0.22	-0.01	0.09	0.19	0.08
SCIENTISTS	0.30	-0.03	-0.14		0.11	-0.37	-0.31	0.13	0.36
SATISFIED	0.09	0.65	0.22	0.11		0.05	0.10	0.23	0.10

DISCUSSION

The assumption underlying deliberative methods is that social preferences can be defined in a group setting in which citizen stakeholders engage deeply with each other and with subject matter experts in order to reveal widely held social values. It is generally assumed, and supported by the limited data, that deliberative group preferences will differ from those that would be obtained by aggregating the preferences of the participating individuals (Irvine et al. 2016, Orchard-Webb et al. 2016). This is because groups may be better able than individuals to form and consider transcendental values such as rights, responsibility, equity, and fairness (Kenter et al. 2016). However, calculative aggregation of individual values has typically occurred by computing the average; other methods of aggregation may compare differently against deliberative group preferences. If some of these methods correlate more closely with the deliberative result, it may tell us something about the way in which groups reach agreement. Further, previous studies have not systematically characterized the degree to which individual preferences may converge as a result of deliberation, or the influences, beliefs, and perceptions associated with such convergence.

We chose our workshop participants to represent a diverse mix of demographic characteristics, including age, sex, income, and political affiliation. Therefore, it is perhaps not surprising that there was substantial disagreement in their individual ranking of ecosystem services before they had a chance to interact with each other or the project scientists. This disagreement is characterized by high within-group standard deviations overall (Table 3) and for most individual ecosystem services (Table 4).

Despite their individual differences, all groups but one were able to reach consensus on all ecosystem service rankings after the full day of engagement. Further, the group rankings they achieved could not, in general, have been revealed by simply aggregating their individual predeliberation rankings. None of the aggregation rules had average Kendall correlation coefficients against the consensus rankings that were greater than 0.3. Notably, however, although Group 1 had individual rankings that were highly discordant (SD = 0.93) (Table 3), the Borda method was able to exactly reproduce their group deliberative ranking. The Kemeny and Copeland methods also came quite close. The fact that the Borda aggregation, a consensus-based method, performed best across all groups in recreating group rankings suggests that the deliberations included elements of consensus building.

Consensus-based aggregation of predeliberation preferences was not enough to reproduce deliberative preferences. This required

some change in individual rankings. Our postdeliberation survey revealed that participants' preferences evolved in the course of interacting with each other and with scientists. Rather than stick with their initial rankings, most participants ranked the 10 ecosystem services in a way that conformed much more closely to their group's ranking and to each other (Table 6). This change is apparently what allowed the deliberative group rankings to be reproduced accurately through aggregation of individual postdeliberation rankings using Kemeny and Copeland methods. For all groups, these two methods achieved Kendall correlations of 0.8 or greater (Table 5). In most cases, the Borda method was not substantially worse. These three methods are distinctive in considering the entire ranked preference list of each participant, leading to an aggregate ranking of broad appeal rather one that focuses on just identifying the ranking most preferred by the majority. This is consistent with the results of Ito et al. (2009), who found that the differences between individual and collective expressions of preference were smaller when collective assessments were made using a consensus rule rather than a majority decision rule.

Convergence in the preferences of group members is demonstrated by a substantial reduction in within-group standard deviations overall (Table 3) and for specific ecosystem services (Table 4). Beliefs about the relative importance or unimportance of Forest Type, Fish Habitat, and Coastal Health were especially convergent within groups (Table 4). Interestingly, however, opinions on these ecosystem services were not always shared across groups, as evidenced by high across-group standard deviations for these three services (Fig. 2). These results hint at the occurrence of groupthink, a phenomenon by which members of a group, in an effort to maintain harmony, reach a consensus decision without critical evaluation of alternative viewpoints. In our context, the lack of creative and independent thought and discourse caused by groupthink has the risk of leading to an irrational or misleading group ranking. On the other hand, for some ecosystem services, such as Flood Protection, Heat Regulation, Snow Cover, and Recreation Days, there was both within- and across-group convergence (Table 4; Fig. 2). This suggests that some preferences may truly have been socially constructed in the group setting in a manner that is largely reproducible. Further, the fact that participants largely maintained the rankings of their group in their individual postdeliberation rankings (Table 6) suggests that they truly internalized their changed preferences rather than simply seeking to conform to the group opinion.

One group was certainly not a victim of groupthink. Members of Group 4 were in disagreement coming into the group deliberation

and only became more discordant afterwards. The within-group standard deviation rose from 0.86 to an unrivaled 0.91 (Table 3). The fact that this was the only group to assign ties to top-ranked ecosystem services in two domains suggests that in the face of such disagreement, they had trouble reaching a clear consensus ranking. Yet, despite their unresolved disagreement, even Group 4's postdeliberation individual rankings could be aggregated using the Borda, Kemeny, or Copeland method to reproduce their group deliberative results. Thus, it seems that even when the preferences of participants did not converge in the process of deliberation, their deliberative ranking still somehow represents the aggregation of their individual rankings. In other words, members of a group could "agree to disagree" and choose a "midpoint" of their opinions to represent this compromise ranking. The Kemeny aggregate ranking, for example, is the one that minimizes the sum of the Kendall tau distances to the individual participants' rankings.

Although many participants changed their rankings over the course of the workshops, our survey results revealed that most were not aware of these changes. In addition to the majority asserting that they only "slightly" changed their opinions, there was a very weak negative correlation with a measure of the consistency between their pre-deliberation and postdeliberation rankings (τ_{ind}). Instead, it seems that participants' self-assessment of change was influenced by how strongly their own ranking converged on the group's ranking over the course of deliberation ($\tau_{post} - \tau_{pre}$) (Table 7). Their actual consistency of ranking from pre-deliberation to postdeliberation is negatively related to the influence they felt the scientists had on the process (τ_{ind} versus SCIENTISTS) (Table 7). The perception of participants about the influence of the scientists was also strongly positively associated with $\tau_{post} - \tau_{pre}$. Together, these associations indicate that the participants whose opinions changed most substantially to reach the group ranking tended to attribute that change to the influence of the scientists.

Even though they were subject to the sometimes uncomfortable feeling of changing their opinion, participants were overwhelmingly satisfied with the outcome of the deliberation. Based on the cross-correlations with other questions, this seems to be because they felt that their opinion was heard and that they had an influence on the outcome. Ito et al. (2009) also found that reduced disparity between individual and collective assessments of willingness to pay for ecosystem services corresponded with increased participant satisfaction with the collective result. Interestingly, in our study, there was only a very weak correlation between the perception of influence and a measure of influence, the correspondence between their initial individual ranking and the group's ranking, τ_{pre} . In actuality, the perception of influence seems to be driven more by how strongly the participant followed rather than led the group's views τ_{pre} .

While addressing some key issues regarding the nature of the deliberative valuation process, our study also raises some additional questions. If group deliberative results can be obtained by aggregating individual rankings after deliberation, then is it necessary for the group to actually go through the process of reaching a consensual ranking? Would the same results be obtained by simply having the participants qualitatively discuss and debate their evaluations and then applying the Kemeny or

Copeland method to their individual postdeliberation rankings? Wilson and Howarth (2002) suggest that a coordinated task activity is an essential element of a successful deliberative valuation process. It would be useful to test this assertion. For cost and convenience, we surveyed our participants after deliberation on the same day. Are the preference changes they experienced maintained over time, or do they eventually revert back to their initial opinions? Answering this question would reveal whether the opportunity the workshops provided for meaningful exchange with other citizen stakeholders and scientists were influential, even revelatory, experiences versus only occasions for swaying short-term stated opinions.

CONCLUSIONS

To conclude, our main findings are that the relative importance citizen stakeholders placed on ecosystem services changed after group deliberation and access to scientists. In general, the ecosystem service rankings of individuals in a group converged toward each other and toward the group deliberative ranking. While the group ranking could not be reproduced from the pre-deliberation individual rankings using any of the aggregation methods considered, it could be largely reproduced from postdeliberation rankings that were aggregated according to the Kemeny and Copeland methods. The defining feature of these two methods is that they comply with the Condorcet criterion, which tends to give a full aggregate ranking of broad appeal rather than either an average ranking or one that emphasizes the choice ranked highest by the majority. This suggests that in reaching their group ranking, participants deliberated and compromised on the importance of each ecosystem service rather than simply averaging their opinions or tallying first-place votes. Postdeliberation survey results suggest that the participants whose preferences changed the most during the deliberation tended to attribute this change to the influence of the scientists. This supports the importance of giving groups access to scientific information as they deliberate on their shared values for ecosystem services.

Responses to this article can be read online at:

<http://www.ecologyandsociety.org/issues/responses.php/9519>

Acknowledgments:

This research was funded by NSF EPSCoR #EPS 1101245.

LITERATURE CITED

- Arrow, K. J. 1951. *Social choice and individual values*. John Wiley & Sons, New York, USA.
- Burgman, M. A., H. M. Regan, L. A. Maguire, M. Colyvan, J. Justus, T. G. Martin, and K. Rothley. 2014. Voting systems for environmental decisions. *Conservation Biology* 28(2):322–332. <http://dx.doi.org/10.1111/cobi.12209>
- Chamberlin, J. R. 1985. An investigation into the relative manipulability of four voting systems. *Systems Research and Behavioral Science* 30(4):195–203. <http://dx.doi.org/10.1002/bs.3830300404>

- Copeland, A. H. 1951. *A reasonable social welfare function*. University of Michigan, Ann Arbor, Michigan, USA.
- García-Llorente, M., B. Martín-López, I. Iniesta-Arandia, C. A. López-Santiago, P. A. Aguilera, and C. Montes. 2012. The role of multi-functionality in social preferences toward semi-arid rural landscapes: an ecosystem service approach. *Environmental Science & Policy* 19:136–146. <http://dx.doi.org/10.1016/j.envsci.2012.01.006>
- Haida, C., J. Rüdiger, and U. Tappeiner. 2016. Ecosystem services in mountain regions: experts' perceptions and research intensity. *Regional Environmental Change* 16(7):1989–2004. <http://dx.doi.org/10.1007/s10113-015-0759-4>
- Hajkowicz, S. 2006. Multi-attributed environmental index construction. *Ecological Economics* 57(1):122–139. <http://dx.doi.org/10.1016/j.ecolecon.2005.03.023>
- Hostmann, M., M. Borsuk, P. Reichert, and B. Truffer. 2005. Stakeholder values in decision support for river rehabilitation. *Archiv für Hydrobiologie. Supplementband. Large Rivers* 15(1–4):491–505. <http://dx.doi.org/10.1127/lr/15/2003/491>
- Howarth, R. B., and M. A. Wilson. 2006. A theoretical approach to deliberative valuation: aggregation by mutual consent. *Land Economics* 82(1):1–16. <http://dx.doi.org/10.3368/le.82.1.1>
- Irvine, K. N., L. O'Brien, N. Ravenscroft, N. Cooper, M. Everard, I. Fazey, M. S. Reed, and J. O. Kenter. 2016. Ecosystem services and the idea of shared values. *Ecosystem Services* 21:184–193. <http://dx.doi.org/10.1016/j.ecoser.2016.07.001>
- Ito, N., K. Takeuchi, K. Kuriyama, Y. Shoji, T. Tsuge, and Y. Mitani. 2009. The influence of decision-making rules on individual preferences for ecological restoration: evidence from an experimental survey. *Ecological Economics* 68(8):2426–2431. <http://dx.doi.org/10.1016/j.ecolecon.2009.03.022>
- Karjalainen, T. P., M. Marttunen, S. Sarkki, and A.-M. Rytönen. 2013. Integrating ecosystem services into environmental impact assessment: an analytic-deliberative approach. *Environmental Impact Assessment Review* 40:54–64. <http://dx.doi.org/10.1016/j.eiar.2012.12.001>
- Keeney, R. L., and H. Raiffa. 1993. *Decisions with multiple objectives: preferences and value tradeoffs*. Cambridge University Press. <http://dx.doi.org/10.1017/CBO9781139174084>
- Kemeny, J. G. 1959. Mathematics without numbers. *Daedalus* 88(4):577–591.
- Kenter, J. O. 2016. Integrating deliberative monetary valuation, systems modelling and participatory mapping to assess shared values of ecosystem services. *Ecosystem Services* 21(Part B):291–307. <http://dx.doi.org/10.1016/j.ecoser.2016.06.010>
- Kenter, J. O., T. Hyde, M. Christie, and I. Fazey. 2011. The importance of deliberation in valuing ecosystem services in developing countries—evidence from the Solomon Islands. *Global Environmental Change* 21(2):505–521. <http://dx.doi.org/10.1016/j.gloenvcha.2011.01.001>
- Kenter, J. O., N. Jobstvogt, V. Watson, K. N. Irvine, M. Christie, and R. Bryce. 2016. The impact of information, value-deliberation and group-based decision-making on values for ecosystem services: integrating deliberative monetary valuation and storytelling. *Ecosystem Services* 21:270–290. <http://dx.doi.org/10.1016/j.ecoser.2016.06.006>
- Kenter, J. O., L. O'Brien, N. Hockley, N. Ravenscroft, I. Fazey, K. N. Irvine, M. S. Reed, M. Christie, E. Brady, R. Bryce, A. Church, N. Cooper, A. Davies, A. Evely, M. Everard, R. Fish, J. A. Fisher, N. Jobstvogt, C. Molloy, J. Orchard-Webb, S. Ranger, M. Ryan, V. Watson, and S. Williams. 2015. What are shared and social values of ecosystems? *Ecological Economics* 111:86–99. <http://dx.doi.org/10.1016/j.ecolecon.2015.01.006>
- Koschke, L., C. Fuerst, S. Frank, and F. Makeschin. 2012. A multi-criteria approach for an integrated land-cover-based assessment of ecosystem services provision to support landscape planning. *Ecological Indicators* 21:54–66. <http://dx.doi.org/10.1016/j.ecolind.2011.12.010>
- Lienhoop, N., and M. Völker. 2016. Preference refinement in deliberative choice experiments for ecosystem service valuation. *Land Economics* 92(3):555–577. <http://dx.doi.org/10.3368/le.92.3.555>
- Liu, S., M. Hurley, K. E. Lowell, A.-B. M. Siddique, A. Diggle, and D. C. Cook. 2011. An integrated decision-support approach in prioritizing risks of non-indigenous species in the face of high uncertainty. *Ecological Economics* 70(11):1924–1930. <http://dx.doi.org/10.1016/j.ecolecon.2011.05.021>
- Mavrommati, G., M. E. Borsuk, and R. B. Howarth. 2016. A novel deliberative multicriteria evaluation approach to ecosystem service valuation. *Ecology and Society* 22(2):39. <http://dx.doi.org/10.5751/ES-09105-220239>
- Meinard, Y., M. Dereniowska, and J.-S. Gharbi. 2016. The ethical stakes in monetary valuation methods for conservation purposes. *Biological Conservation* 199:67–74. <http://dx.doi.org/10.1016/j.biocon.2016.04.030>
- Orchard-Webb, J., J. O. Kenter, R. Bryce, and A. Church. 2016. Deliberative democratic monetary valuation to implement the ecosystem approach. *Ecosystem Services* 21:308–318. <http://dx.doi.org/10.1016/j.ecoser.2016.09.005>
- Proctor, W., and M. Drechsler. 2006. Deliberative multicriteria evaluation. *Environment and Planning C: Politics and Space* 24(2):169. <http://dx.doi.org/10.1068/c22>
- Robbins, L. 1938. Interpersonal comparisons of utility: a comment. *Economic Journal* 48:635–641. <http://dx.doi.org/10.2307/2225051>
- Saaty, T. L. 1980. *The analytic hierarchy process*. McGraw-Hill, New York, USA.
- Samal, N. R., W. Wollheim, S. Zuidema, R. Stewart, Z. Zhou, M. M. Mineau, M. Borsuk, K. H. Gardner, S. Glidden, T. Huang, D. Lutz, G. Mavrommati, A. M. Thorn, C. P. Wake, and M. Huber. 2017. A coupled terrestrial and aquatic biogeophysical model of the Upper Merrimack River watershed, New Hampshire, to inform ecosystem services evaluation and management under climate and land-cover change. *Ecology and Society* 22(4):18. <http://dx.doi.org/10.5751/ES-09662-220418>
- Schuwirth, N., P. Reichert, and J. Lienert. 2012. Methodological aspects of multi-criteria decision analysis for policy support: a

case study on pharmaceutical removal from hospital wastewater. *European Journal of Operational Research* 220(2):472–483. <http://dx.doi.org/10.1016/j.ejor.2012.01.055>

Stagl, S. 2006. Multicriteria evaluation and public participation: the case of UK energy policy. *Land Use Policy* 23(1):53–62. <http://dx.doi.org/10.1016/j.landusepol.2004.08.007>

TEEB. 2010. The economics of ecosystems and biodiversity: ecological and economic foundations. P. Kumar, editor. Earthscan, London, UK.

Wegner, G., and U. Pascual. 2011. Cost-benefit analysis in the context of ecosystem services for human well-being: a multidisciplinary critique. *Global Environmental Change* 21(2):492–504. <http://dx.doi.org/10.1016/j.gloenvcha.2010.12.008>

Wilson, M. A., and R. B. Howarth. 2002. Discourse-based valuation of ecosystem services: establishing fair outcomes through group deliberation. *Ecological Economics* 41(3):431–443. [http://dx.doi.org/10.1016/S0921-8009\(02\)00092-7](http://dx.doi.org/10.1016/S0921-8009(02)00092-7)

Zia, A., P. Hirsch, H. Van Thang, T. C. Trung, S. O'Connor, T. McShane, P. Brosius, and B. Norton. 2015. Eliciting intertemporal value trade-offs: a deliberative multi-criteria analysis of Vietnam's Bai Tu Long National Park management scenarios. *Journal of Sustainability, Energy and the Environment* 2(1):41–62. <http://dx.doi.org/10.22492/ijsee.2.1.03>

Zoderer, B. M., P. S. L. Stanghellini, E. Tasser, J. Walde, H. Wieser, and U. Tappeiner. 2016. Exploring socio-cultural values of ecosystem service categories in the Central Alps: the influence of socio-demographic factors and landscape type. *Regional Environmental Change* 16(7):2033–2044. <http://dx.doi.org/10.1007/s10113-015-0922-y>

Zografos, C., and R. B. Howarth. 2010. Deliberative ecological economics for sustainability governance. *Sustainability* 2(11):3399–3417. <http://dx.doi.org/10.3390/su2113399>